ESD-TDR-64-254

SHORT-TERM MEMORY: NON-EQUIVALENCE OF QUERY AND MESSAGE ITEMS

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-64-254

FEBRUARY 1964

James D. Baker
Walter E. Organist

DECISION SCIENCES LABORATORY
DEPUTY FOR ENGINEERING AND TECHNOLOGY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 4690, Task 469003

DDC AVAILABILITY NOTICES

ESD-TDR-64-254

SHORT-TERM MEMORY: NON-EQUIVALENCE OF QUERY AND MESSAGE ITEMS

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-64-254

FEBRUARY 1964

James D. Baker
Walter E. Organist

DECISION SCIENCES LABORATORY
DEPUTY FOR ENGINEERING AND TECHNOLOGY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 4690, Task 469003

FOREWORD

The subjects for this study were provided by the Regis College for Women, Weston, Massachusetts, under Decision Sciences Laboratory (DSL) Contract AF 19(604)-5958. The assistance of Arthur Marcus and the constructive comments of Robert Westfield and William Sumby of DSL and Douwe B. Yntema of M.I.T.'s Lincoln Laboratory are gratefully acknowledged.

ESD-TDR-64-254

SHORT-TERM MEMORY: NON-EQUIVALENCE OF QUERY AND MESSAGE ITEMS

ABSTRACT

This study was designed to discover whether query and message items are equivalent in their effect on short-term memory in situations where a person is required to process a sequence of messages while concurrently processing queries about them. It was assumed that recall is consistently degraded as the number of items interpolated between a message and its subsequent query increases. It was hypothesized that this degradation is greater for those subsequences with interpolated queries than for those containing messages only. The results showed that interpolated queries do degrade recall significantly more than do messages.

PUBLICATION REVIEW AND APPROVAL

This Technical Documentary Report has been reviewed and is approved.

Donald W. Connolly

DONALD W. CONNOLLY
Chief, Display Division
Decision Sciences Laboratory

Roy Morgan

ROY MORGAN, Colonel, USAF
Director
Decision Sciences Laboratory

KEY WORD LIST

1. HUMAN ENGINEERING
2. MEMORY
3. BEHAVIOR
4. PSYCHOLOGY
5. EXPERIMENTAL DATA
6. PERCEPTION

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Experiments in human memory usually are concerned with recall either of data associated with recent, once experienced events (short-term memory) or of data established by repetitive learning experiences (long-term memory). The study reported here is concerned with short-term memory (STM).

The classical approach to the study of STM has generally utilized a memory-span paradigm, i.e., the presentation of items to be remembered and the recall of these items are discretely separated. Outside of the laboratory, however, STM is often continuous in nature. Many everyday activities require that a person recall, at random intervals, information which he has been storing while he continues to process and store new information as, for example, in playing cards. Consequently, recent experimentation, e.g., Yntema & Mueser (1960) and Lloyd, Reid & Feallock (1960), has begun to examine STM in the context of continuous tasks.

Basic to this class of experiment is a stimulus list consisting of a random sequence of message and query items. A test session begins with the presentation to the subject (S) of the initial stimulus, which is necessarily a message item. A query about that item may then follow immediately in the sequence, or it may be separated from the appropriate message by other messages, by other queries, or by both. Once the sequence has begun, S must continuously store messages and concurrently process queries about them.

A usual finding in studies of this type is that performance is consistently degraded by increasing the number of items interpolated between the presentation of a stimulus and its recall, although performance remains above chance

(Posner, 1963). No distinction is made between query and message as an intervening item. There are, however, reasons for believing such a distinction to be important.

If humans process information in the same manner as do machines, then the information processing steps for a message would be input, process and store in memory. The information processing steps for a query would be input, process, search memory, retrieve and output. Since, assuming only one processing step can occur at a time, perceptual organization and rehearsal must occur *between* items in a continuous STM task, the extra steps required in processing a query should result in less opportunity for intervening organization and rehearsal than does a message, especially if the task is paced. Therefore, an intervening query should degrade recall more than does a message, and performance should be affected by type as well as number of items interpolated between the presentation of a stimulus and its recall. The following study was conducted to test the hypothesis that performance is degraded more by the interpolation of queries than of messages.

METHOD

Stimulus Lists

Figure 1 presents schematically the elements from which message items were formed. Each of the twenty-four possible combinations of object, attribute and state was drawn four times to generate a list of ninety-six randomly ordered message items (e.g., "A-red"). Three queries for

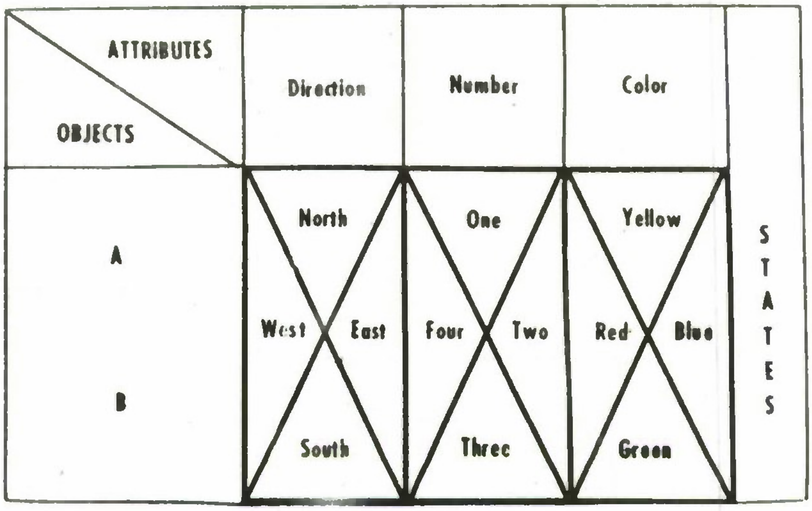

Figure 1. Population of message items used. Examples of single message items would be: A-red; A-north; A-three - likewise, B-red; B-north; etc.

---

each attribute (color, direction, number) were inserted within this list - a total of nine queries (e.g., "Color of A?"). For each attribute, one query followed the message item bearing the correct answer by a lag of five intervening items, one query by a lag of seven items, and one by a lag of nine. A lag is defined as the number of items from a given query back to and including the message item containing the correct answer. Some rearrangement within the random list was necessary to fulfill this design. The result was a quasi-random list (sublist A) which had approximately equal distribution of objects, attributes, states, lags and attribute queries, with *no intervening queries* as lag items.

Queries referring to the immediately preceding message item ("no lag" queries, or, to be consistent, queries with a lag of one) were substituted for the midpoint message item of each lag in sublist A to produce sublist B.

Sublist A	*Sublist B*
A-red	A-red---------------item 1 (correct response)
A-one	B-one---------------item 2
B-green	Number of B?-------item 3 (intervening query in sublist B)
A-four	A-four--------------item 4
B-north	B-north-------------item 5
Color of A?	Color of A?---------Query with lag of 5

Figure 2. Changing a segment of the basic list by inserting an intervening query for a lag of 5 sequence. Note that the message item which determines the correct response is computed as part of the lag.

Further internal rearrangements were made in sublist A as well as sublist B to meet the constraints that these intervening, "no lag" queries should refer equally to each attribute, and, that an intervening query should not be the same as the next query. The two resulting sublists differed only in that sublist B provided the simplest case of a query as an intervening item. This relationship is shown in Figure 2.

Two equivalent, experimental stimulus lists were then generated by combining the sublists as forms AB and BA. Using either form S was required to keep current in memory the present state of six variables (3 attributes times 2 objects: See Fig. 1). The message and query items were recorded in sequence on audio-tape at 5 second intervals. The inter-item time available for perceptual organization and rehearsal was thus constant throughout the two forms. The end-product was two paced stimulus recordings whose counterbalanced design allowed each S to serve as his own control.

Subjects

Twenty female college students were used in this experiment. None of the S's had experience in this type of study.

Apparatus

The stimulus list was presented using a DeJur/Grundig Stenorette-TD tape recorder (model 50-187) with an auxiliary speaker (model DS-518). The spoken response was recorded using a second Stenorette. The S responded through a small, hand-held microphone which was "open" at all times. Because both recorders were started simultaneously, and the response mike held open, the stimulus list, with the S's responses intervening, was recorded.

Procedure

At the beginning of each experimental session S was handed a set of specially developed programed instructions which oriented her to the task. This programed instruction approach was used to reduce to a minimum the experimenter/subject interaction which could introduce bias into the study and to standardize the pre-data collection treatment of the Ss (Baker, 1963). The instructions also encouraged S to guess. S proceeded through the instructions at her own pace. During this time, the experimenter (E) placed the appropriate tape on the machine for the condition AB or BA to which S had been randomly assigned. The last frame of S's programed instructions stated that she should tell E to turn on the equipment. E then started both tape recorders. The initial portion of the stimulus tape contained supplementary procedural instructions followed by a short practice session. Without interrupting the

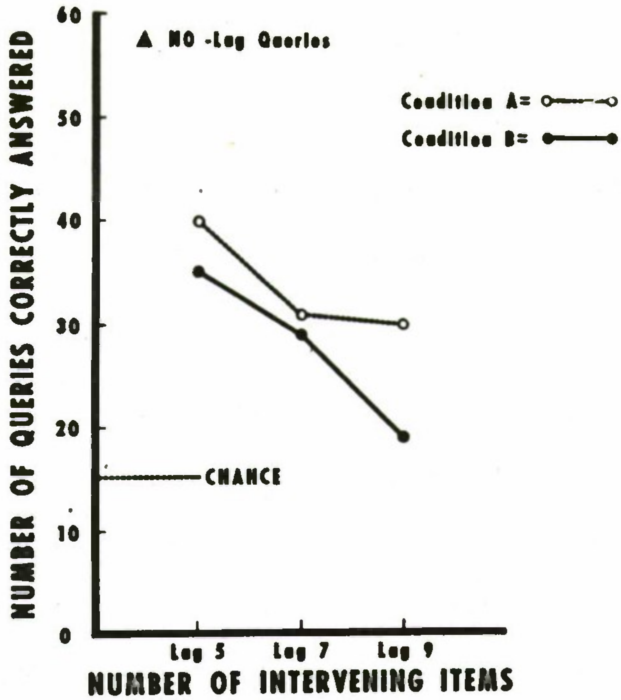

Figure 3. The number of non-homogenous queries correctly answered as a function of lag. Under both conditions S had one query about each of the three attributes for every lag. The 20 Ss served under both condition, hence 3 x 20 = 60 was the total possible number of correct responses for each lag of either condition.

---

play of the tape a transition was effected by a pre-recorded statement that data collection would now begin. The tape then phased directly into the stimulus materials. Throughout the stimulus tape, S was given no feedback. The entire experimental session took approximately one hour per S.

RESULTS

Non-homogenous Interpolated Query

The performance measure in this study was the number of queries (excluding "no lag") answered correctly. The number of correct answers possible was nine per S for each experimental condition. The first analysis tested whether scores under the two conditions were influenced by position in the counterbalanced design, to disclose whether the effects of extraneous variables, e.g., fatigue and practice, were equal for both conditions. A Sign Test comparison was made of the number of queries answered correctly when

Condition A was first and when it was second; the same comparison was made for Condition B. The differences obtained were not statistically significant.

An analysis was next undertaken of the relationship between frequency of correct responses and lag. The results illustrated in Figure 3 show that frequency of correct recall is a negative function of lag for both experimental conditions. Friedman's two-way analysis of variance showed the differences between the number of queries per lag answered correctly to be statistically significant ($p < .02$). This is in keeping with the findings of similar studies, reviewed by Posner (1963), which have shown that performance is consistently degraded by increasing the *number* of items interpolated between the presentation of a stimulus and its recall.

The purpose of the present study, however, was to discover whether degradation of performance is also related to the *type* of item interpolated between the presentation of the stimulus and its recall. It was hypothesized that an interpolated query degrades recall more than does a message. The plots in Fig. 3 support this hypothesis since the level of performance for the intervening query condition (B) is consistently lower than that of the message condition (A). Student's t was used to test for statistical significance of this difference. Queries were paired (N=9 per condition) and the difference between number of correct recalls possible (N=20 for each query) was taken as the basic measure. The difference in performance levels was found to be statistically significant, ($t = 2.65$; $p < .05$). Thus, it may be concluded that interpolated queries do degrade recall more than do messages.

A special case of the effect of elapsed time between items was also noted. Plotted in Fig. 3 is the number of correct responses to "no lag" queries, i.e.,

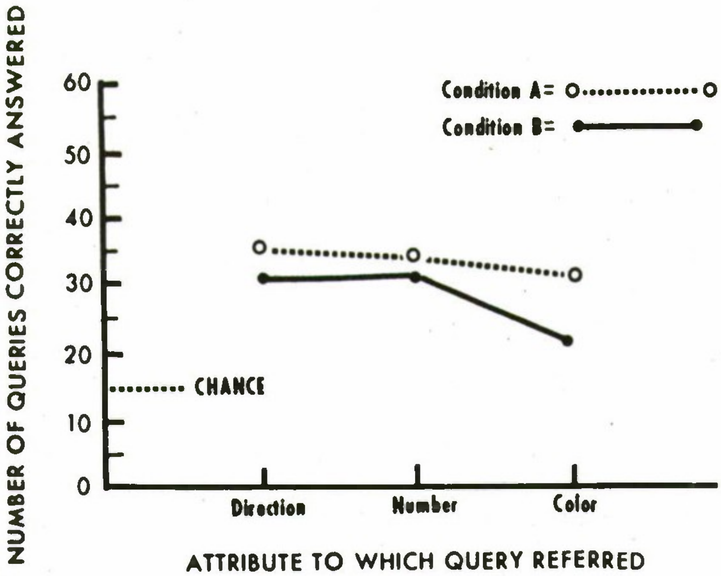

Figure 4. The number of queries answered correctly as a function of the attribute to which the query referred. Each S was presented with 3 queries concerned with each attribute. The N per condition was 20, so total responses per attribute was 60.

---

the midpoint queries which referred to the immediately preceding item. It was found tha a mere five second delay in this paced task prevents perfect recall.

The previously delineated stimulus list constraints also permitted, under controlled conditions, a test of whether recall is affected by the attribute to which the query referred. The results of some recent experiments by Mackworth (1963) suggested that recall in this task may be poorer for color items than for number items. Data illustrating the effect of attribute on recall are presented in Figure 4. An intra-condition analysis (Wilcoxon Test) comparing frequency of correct recalls as a function of attribute showed no statistically significant difference. In Fig. 4 the most obvious difference continues to be in the effect on recall of query as an intervening item.

Sublist B	*Sublist C*
A-red	A-red---------------item 1 (correct response
B-one	B-green-------------item 2
Number of B?	Color of B?---------item 3 (intervening query)
A-four	A-four--------------item 4
B-north	B-north-------------item 5
Color of A?	Color of A?---------Query with lag of 5

Figure 5. The intervening query for sublist B is non-homogenous, i.e., it refers to a different attribute than the lag query which follows it. The matched segment of sublist C has a homogenous intervening query, i.e., both queries, no lag and lag, refer to the attribute *color*.

Homogenous Interpolated Query

The finding that interpolated queries significantly affect recall must be qualified. One of the design constraints in the experiment was that an intervening query should not be of the same class as the next query, i.e., only non-homogenous queries intervened, but all classes of queries were used as intervening items. The constraint was imposed because of the possibility of differential effects of interpolated queries of the same class which could contaminate a simple test of query as an intervening item. Therefore, an additional experiment was conducted to test this possibility.

Sublist B of the previous experiment served as the point of departure for developing the stimulus list for this experiment. The intervening no lag queries, and their immediately preceding message item, were rearranged so that the no lag queries now referred to the same attribute, but not state, as the lag query which followed it. This relationship is illustrated in Figure 5.

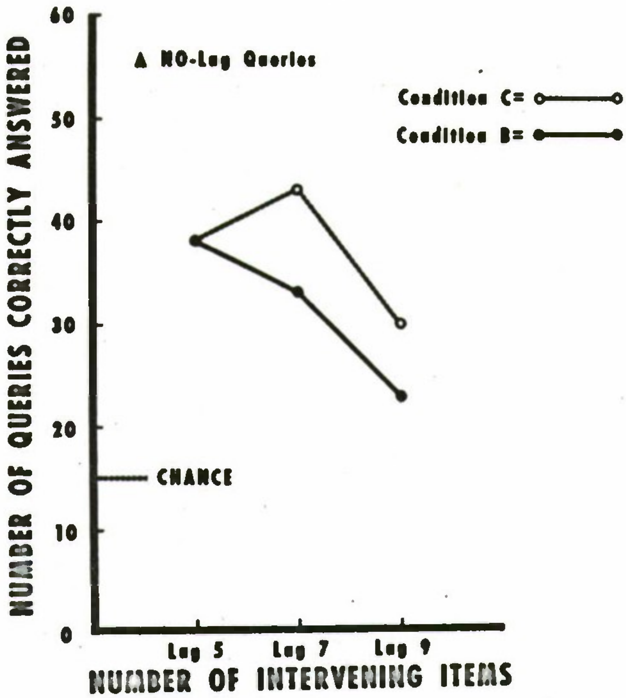

Figure 6. The number of homogenous queries correctly answered as a function of lag. Under both conditions S had one query about each of three attributes for every lag. The 20 Ss served under both conditions, hence 3 x 20 = 60 was the total possible number of correct responses for each lag of either condition.

The result was sublist C. The sublists were then combined into forms BC and CB. The remainder of the design and conduct of this experiment was identical to that of the previous one. The Ss consisted of twenty female college students, none of whom were previous participants.

As in the previous experiment, the possible number of lag queries which could be answered correctly was nine per S for each condition. Differences in the number of queries answered correctly as a function of position in the counterbalanced order (i.e., Condition B first vs. B second; likewise C first vs. C second) were not found to be statistically significant.

Results illustrated in Figure 6 show the frequency of correct recall as a function of lag. With one exception, these results agree with the previous

findings that recall is consistently degraded by increasing the number of items intervening between a query and its appropriate message. The exception was a lag of seven. In this case, intervening homogenous queries appear to degrade recall at the same level as a lag of five. Detailed analysis of the responses to these lag items failed to produce an adequate explanation for this trend. Nevertheless, differences between the number of queries per lag answered correctly were found, overall, to be statistically significant ($p < .02$; Friedman's two-way analysis of variance).

It was hypothesized that homogenous intervening queries may have a less degrading effect on recall than non-homogenous intervening queries. Although a trend in that direction is evident in Fig. 6, Student's *t* test (queries matched between conditions) indicated the mean difference not to be statistically significant. Therefore, there is not sufficient evidence to conclude that homogenous intervening queries differ significantly from non-homogenous intervening queries in their effect on recall.

DISCUSSION

One approach to studying human short-term memory (STM), in the absence of a general theory of memory, is to employ task - specific research models abstracted from "real world" situations.[1] A problem inherent in this approach, however, is that variables embedded within the task itself may

---

[1] For example, in describing the general nature of his experiments in STM, Yntema (1963) has stated: "The sort of situation that we were attempting to abstract and bring into the laboratory is that in which an operator must keep track of several things at once. For example, the information needed by an air-traffic controller. . . . . "

mask or enhance the effects of independent variables, unless the contribution of task variables is known and controlled. The present set of experiments provides information regarding the contribution of some of the task variables common to studies of continuous STM.

For example, continuous STM tasks generally do not use classical experimental stimuli, i.e., nonsense syllables or strings of digits. Rather, the messages to be remembered consist of mutually exclusive sets of every-day words. Thus number, direction and color, as attributes of objects _A_ and _B_, are mutually exclusive since neither the number nor the direction of _A_ can be red. The results of some recent experiments by Mackworth (1963) had suggested that recall in this type of task may be poorer for color items than for number items. However, we did not find this task variable to have a significant effect on recall within the context of the present study.

Also inherent in continuous STM studies are varying time-delays between the presentation of a message and a query about that message, even if the task is paced. The nature of the task is such that for some queries the immediately preceding item may contain the correct response, while in other instances _S_ may need to search far back into memory. We found that STM is impaired simply by the passage of time with no items intervening, i.e., a mere five second delay for the "no lag" items produced a decrement in recall. In those instances where _S_ must search further back into memory, the effect of time - delay is compounded by the processing and storage of intervening items. Since it is more parsimonious in continuous STM studies to address oneself to the problem of effects of intervening items than to tease out the contaminated effects of time-delay, results are usually reported in terms of correct response as a function of number of items in store, number of intervening items, etc., without regard to time-delay.

A usual finding is that performance is consistently degraded by increasing the number of items interpolated between the presentation of a stimulus and its recall (Posner 1963). Our results are in agreement with previous findings, since we found that recall was consistently degraded when 5, 7 and 9 items intervened between a query and its appropriate message.

However, the interpolated items in a continuous STM study are necessarily of two types, viz., queries and messages. It was hypothesized that queries are not equivalent to messages in their effect on recall. We found that intervening queries degrade recall significantly more than do messages. Thus, recall is effected by type (message or query), as well as number, of items intervening between a message and its query.

Related to this are the findings recently reported by Reid, Brackett, and Johnson (1963), although their study was considerably different from the present study. In an extensive examination of several aspects of continuous STM, they found sub-sequences in their stimulus lists which allowed them to assess the impact of interpolated recall points on previously presented items. Although their interpretation of their data was only suggestive, the implications were that specific instances of improved performance were related to a decrease in the number of interpolated recall points that occurred in sub-sequences of the stimulus lists.

It should be made explicit that the results obtained in the present study were found when the relationships of the variables under consideration were structured, while these variables would be randomly distributed over conditions in most continuous STM experiments. Thus, one immediate effect of this study is to point out the criticality of insuring that intervening queries are

evenly distributed over conditions when using an experimental task of this type. In terms of the implications for further research, the question arises as to whether the differential effect of a query as an interpolated item is best interpreted as: (1) an intrusion-interference effect or (2) an erasure expectancy effect. That is, an interpolated query could, like an interpolated message item, interfere with recall but to a greater degree; or, on the other hand, it could serve as a signal for the subject to erase or clear his STM on the expectation that a new "trial" is starting. A second experimental question has to do with items having number as an attribute. The data from both of the experiments reported here strongly suggest that recall was insensitive to effects of experimental conditions when the queries referred to message items which had number as an attribute.

Aside from the methodological and theoretical implications of the experimental results, a practical implication of the present findings should be noted. In computer-based information-processing systems, operators are often required to store in STM data concerning relatively short sequences of events. These data provide the basis for filtering and shunting decisions on the part of the man in the system. Disruption of this memory store, logically, results in less than optimal decision. Hence, one recommendation by the Winter Study Group[1] was that operators should not be required to perform routine, menial tasks in combination with more complex or important tasks if operator performance is to be optimized. An example

[1] The Winter Study Group was a group of experts from government and industry convened by the U. S. Air Force in 1960, to determine the nature of required command and control systems and to advise on the best methods for obtaining them.

given was that operators should not be required to respond to phoned or verbal requests for information when the requests were of low priority and could await a more opportune moment. In the light of our findings, this recommendation is extremely sound since STM would be more suspectible to degradation by phoned or verbal requests (queries) than by handling more data (messages) about the event at hand.

SUMMARY AND CONCLUSIONS

This study was concerned with human short-term memory (STM) in the context of continuous tasks. It was hypothesized that an intervening query in a sequentially presented STM stimulus list degrades recall more than does a message. Two experiments were conducted to test this hypothesis. The first experiment tested for the degrading effects of non-homogenous intervening queries, i.e., the intervening query did not refer to the same class of message as the query that followed it. The findings showed that an intervening query significantly degrades recall more than an intervening message. The second experiment tested for possible less degrading effects on recall when the intervening query is homogenous. No statistically significant difference between homogenous and non-homogenous intervening queries was found. It is concluded that query and message items are not equivalent in their effect on recall and that degradation in recall is a function of both number and type (message or query) of items interpolated between the presentation of a stimulus and its recall.

REFERENCES

BAKER, J. D. Programmed instruction as a methodological tool in psychological research, J. programed Instruction, 1963, 2, 19-23.

LLOYD, K. E., REID, L. S., and FEALLOCK, J. B. Short-term retention as a function of the average number of items presented. J. exp. Psychol., 1960, 60, 201-07.

MACKWORTH, J. F. The relation between the visual image and post-perceptual immediate memory. J. verbal Learn. verbal Behav., 1963, 2, 75-85.

POSNER, M. I. Immediate memory in sequential tasks. Psychol. Bull., 1963, 60, 333-49.

REID, L. S., BRACKETT, H. R., and JOHNSON, R. B. The influence of relationships among items to be recalled upon short-term retention. J. verbal Learn. verbal Behav., 1963, 2, 86-92.

SIEGEL, S. Non parametric statistics for the behavioral sciences. New York: McGraw Hill Book Company, Inc. 1956.

WALKER H. M. and LEV, J. Statistical Inference. New York: Henry Holt & Co., 1953.

YNTEMA, D. B. Keeping track of several things at once. Human Factors, 1963, 5, 7-17.

YNTEMA, D. B. and MUESER, G. E. Remembering the present states of a number of variables. J. exp. Psychol., 1960, 60, 18-22.